Appendix B

Guidelines for data development for key public health problems

Data advisory panel members:

Tom Bell, DOH, Community and Family Health
Sherilynn Casey, DOH, the PHIP Activity Standards Committee
Joe Castorina, DOH, Community and Family Health
Sharon Estee, DOH, Epidemiology and Health Statistics
Diane Gordon, Snohomish County Health Department
Chris Hale, Tacoma-Pierce Health Department
Steve Kelso, DOH, Office of Policy and Planning
John Kobayashi, DOH, Epidemiology and Health Statistics
Mary LeMier, DOH, Community and Family Health
Steve Macdonald, University of Washington
Lin Song, Seattle-King Health Department
Juliet VanEenwyk, DOH, Epidemiology and Health Statistics, Chair
Frank Westrum, DOH, Environmental Health

Preface

In February 1994, Bobbie Berkowitz, chair of the Steering Committee of the Public Health Improvement Plan (PHIP), convened a Data Advisory Panel (panel) whose short term task was to develop guidelines for the generation and presentation of data in the PHIP. A written summary of the guidelines was reviewed by panel members in May. The guidelines were revised and submitted to reviewers at the University of Washington and Centers for Disease Control and Prevention in July. The following report includes the guidelines for data generation and the rationale for these recommendations.

A summary of the guidelines, which does not include the rationale, is included as Attachment 1. Readers, who do not feel the necessity of understanding the complete rationale for the proposed guidelines, are strongly urged to read the introduction that follows, before turning to Attachment 1. The introduction provides the context for interpretation of data generated using the guidelines.

Introduction

"The art of epidemiologic reasoning is to draw sensible conclusions from imperfect data." (Anonymous) The recognition of the imprecision of epidemiologic measurement applies to data generation as well as data interpretation. This imprecision makes it impossible to prescribe approaches to data generation which are universally applicable. Therefore, the following recommendations are intended as guidelines for data related to Key Public Health Problems in the PHIP.

The primary purposes of the guidelines are to assure that data in the PHIP are 1) generated in a scientifically reasonable manner, 2) presented in a manner which is consistent with standard requirements for displaying scientific material, and 3) useful to Washington State Department of Health (DOH), Local Health Departments and others interested in using the PHIP. The guidelines are not intended to be used in a cookbook fashion. When using data for assessment and policy development, standardized approaches to data can not substitute for epidemiologic and programmatic experience.

The panel recommends the use of statistical procedures as an aid in the interpretation of data. The panel recognizes that the use of statistics with population data is controversial and that in many instances, there are violations of the underlying assumptions (e.g., independence) on which the statistical procedures are based. Therefore, in this context, statistical testing is not intended as formal hypothesis testing and does not substitute for thoughtful attention to the many factors which need to be considered when using data to make public health decisions.

Given this perspective on the use of statistical testing, the panel did not address the issue of multiple comparisons. However, those interpreting the data must be aware that with an alpha of 0.05 or a 95% confidence interval, five out of 100 times, one will see "statistically significant" differences by chance alone, i.e. when there is actually no difference between two data points. Additionally, when there are relatively few events, the statistical power to detect real differences is limited and the violation of assumptions becomes more critical.

Development of baselines, outcome standards and thresholds

The recommendations for the development of baselines, outcome standards and thresholds are applicable to the health and risk status indicators. However, the guidelines may also be used for other types of quantitative indicators. They are not applicable to "qualitative" indicators. The panel recommends that qualitative indicators be made quantitative whenever possible.

Baseline data

In Webster's New Riverside University Dictionary (New Riverside Publishing Co., 1988), baseline is defined as a "line serving as a base, as for measurement." A "base" is .."an observation .. from which a ... process is begun." While the term, baseline, has been adopted by researchers often to indicate the point where one is when a process begins, the original concept of a line focuses on the importance of looking at a number of points over time to determine where one is now.

Generalconsiderations

Use of Rates: For most indicators, the baseline should be expressed as a rate (i.e. number of events per population in a specified time period). Rates are not recommended when small numbers of events serve as sentinel indicators (red flag) of public health concern, e.g., cases of polio. The panel also recommends avoiding the calculation of rates when there are fewer than five events, whether or not the event is viewed as a sentinel indicator. Since estimates may vary due to using different denominators, to ensure consistency, it is recommended that rates be calculated with population denominators provided by the Washington State Center for Health Statistics.

Use of Point Estimates: The panel considered the relative merits of providing point estimates for the baseline statistic versus providing ranges by including confidence intervals (CIs) around the point estimate. The panel opted for the point estimate as a more straightforward manner of providing Washington State data to non-technical readers of the PHIP.

Age-adjustment: There are not hard and fast rules to determine when age-adjustment and use of stratum-specific (e.g., age-, race- and sex-specific) data are appropriate. However, the panel recommends:

- *If there are comparable Healthy People 2000 (U.S. Public Health Service, DHHS Publication No. (PHS) 91-50212, 1991) baselines, rates should be generated in a similar manner relative to age-adjustment and group specificity.*
- *If adjustment is appropriate, data should be adjusted to the 1940 U.S. standard population, except for cancer data. Cancer incidence should be adjusted to the 1970 U.S. standard population. Cancer mortality should be adjusted to both the 1940 and 1970 standards. Although there is debate over whether these standards should continue to be used, these are the standards in Healthy People 2000 and in publications from the National Center for Health Statistics and the National Cancer Institute. For comparative purposes, it is important to be consistent with national standards.*
- *Standard methods for adjustment are presented in most introductory epidemiology texts. In most instances, direct age-adjustment will be used (i.e. age-specific rates in the population of interest will be applied to a standardized age distribution.) Methods of age-adjustment in Healthy People 2000 are presented in Appendix II of Health United States 1992 (DHHS Pub. No.(PHS)93-1232) and Health United States 1993 (DHHS Pub. No.(PHS)94-1232).*
- *For events which increase with age, age-adjustment to the relatively young 1940 and 1970 U.S. standard populations can obscure the magnitude of a problem. Therefore, the number of events should be reported in addition to the adjusted rates.*

Calculation of baselines

Single year baselines

When possible, the single, most recent year of data should be used for the baseline statistic. It is generally possible to use a single year when historic data indicate an underlying trend without substantial annual fluctuation. The determination of whether there is substantial annual fluctuation is both a quantitative and qualitative process.

There are several quantitative approaches. The most standard quantitative approach is to conduct a statistical test for the difference between two rates (proportions). Formulas for these calculations can be found in standard biostatistics and epidemiology texts. If the test indicates a statistically significant difference (e.g., $p<0.05$) between the rates for two consecutive years, one would conclude that there is substantial year to year fluctuation. This approach is similar, although not identical, to setting CIs around point estimates for two adjacent years. If the CIs do not overlap, one would conclude that there is substantial fluctuation.

As indicated in the introduction, in the context of the PHIP, the panel views statistical tests as aids for interpreting data. Qualitative judgements are important for the final determination of whether a single year represents the true situation at baseline. These judgements must be based on a knowledge of what the statistic denotes and a knowledge of the program area to which the data pertain.

There are several factors which need to be remembered when interpreting the results of statistical tests.

- *Because the result of a statistical test depends as much on the number of events as it does on the magnitude of the difference between two rates, there may be instances where the point estimates between adjacent years are not statistically significantly different, but a single year does not represent a stable baseline (i.e., when the number of events is small, there is low power to detect real differences). Therefore, one must be cautious in selecting a single year of data for the baseline when there are fewer than 100 events per year.*
- *When performing a statistical test many times, there is increasing likelihood of finding what appears to be a statistically significant difference, when in reality, there is no difference.*
- *The panel discourages the practice of using an alpha of less than or equal to 0.05 (i.e. $p<=0.05$) as a rigid cut point to determine statistical significance. If $p<0.05$, two points have a less than 5 in 100 probability of being that different by chance when they are really the same. A $p=0.06$ means that there is a 6 in 100 chance of the points being that different, when they are really the same. The panel believes that it is arbitrary to conclude that in the former case, chance is not operating and the points are different, but in the latter case, chance may be operating and so the points are the same.*
- *If the shape of the underlying distribution is not as assumed or if the rate for one year is not independent of the rate for an adjacent year, the statistical test can err in either direction (i.e., finding differences where none exist or failing to find differences where they do exist).*

Qualitatively, because of annual fluctuation, the true situation at baseline falls within a range of values. If the point estimate for a given year is at the low or high end of the range, it may not be a good representation for the baseline. Additionally, knowledge of *changes in reporting methods, coding standards and other events needs to be brought to bear on the decision of whether the rate for a single year is an adequate baseline statistic.*

Consider the following data for mortality from fires and burns.

Deaths from unintentional fire and burns

	Number	Rate/100,000	95% Poisson CIs
1989	55	1.18	0.89-1.53
1990	67	1.38	1.07-1.75
1991	44	0.88	0.64-1.18

Source: Department of Health, Annual Vital Statistics Reports for 1989-1991 ICD9 codes: E890-899, E929.4

All three years of data in this example show overlapping CIs and thus, it may be reasonable to conclude that the rate is stable enough to use the rate for 1991 as the baseline. The relatively small number of events, however, leads to low power to detect real differences. Qualitatively, unless something specifically influenced the rate for 1991, it seems that the rate may be speciously low and, therefore, not adequate as a baseline point estimate. In this example, additional years of data might help to resolve whether the rate for 1991 is a good indicator of where one is at baseline.

Judgement and experience must be brought to bear on the consideration of whether something has occurred which might influence a rate for a given year. For example, one large fire in which many people died could inflate the rate for a particular year. In this example, if there had been a statewide intervention to promote the use of smoke detectors toward the end of 1990, the 1991 rate may be an indication of the success of the program and a reliable indicator to use as a baseline statistic.

If a single year is not adequate for the baseline statistic, the panel recommends two methods of calculating a baseline. One can use moving (rolling) averages to see if stable data points can be generated or one can use a regression methodology. Because the averaging method is likely to be more understandable to non-technical readers of the PHIP, it is recommended that this method be tried first.

Moving-average baselines

Moving averages reduce variability between rates for adjacent time periods by incorporating a given point estimate in several adjacent periods. The panel recommends using three-year moving averages. Therefore, each annual data point contributes to three adjacent time periods. Two-year rolling averages may not be sufficient to smooth the data. If more years are used, it becomes more difficult to measure annual change. Even with three years, change for a subsequent year may be obscured since the new "data point" will be weighted toward the two previous years.

When three-year moving averages are used, adjacent data points contain two years of identical data. Given this extreme lack of independence, the statistical tests described for single years of data cannot be used to inform the decision of whether the most recent three year average is a relatively accurate representation of the situation at baseline. Those generating the data will need to make the decision, in a relatively subjective *manner, based primarily on the shape of the curve. If it is decided that there continues to be substantial fluctuation between adjacent data points, the panel recommends using the regression method described below.*

Regression baselines

A regression line can be fitted to historic data and the regression point for the most recent year of historic data can be used as the baseline. The panel recommends fitting the historic data to several models in order to decide which model provides the best fit. This approach requires a sufficient number of annual data points and the panel does not recommend this approach if there are less than five years of data.

Calculation of outcome standards

The panel defined an outcome standard as an objective which one wishes to accomplish. In the PHIP, the outcome standards refer to goals or targets for the year 2000. Since an outcome standard is a goal, the panel recommends that unless a condition is biologically linked to a specific racial/ethnic group, there should not be separate outcome standards for different groups. This does not mean that one should not look at where different sub-groups are in relation to the goal and assess whether the sub-group has exceeded a threshold. (See section on thresholds.) It simply means that the goals for all groups should be the same even though it may not be possible for every group to meet the target in the same time frame. Groups who are already doing better than the target should at least try to maintain their current level.

The panel developed guidelines for calculating outcome standards when 1) baseline data and year 2000 targets are available at the national level and 2) there are no national goals.

When baseline data and year 2000 targets are available at the national level, the relationship of Washington baseline data to the national figures falls into one of three categories.

1. *The Washington baseline represents a situation which is worse than the national baseline (i.e. it is higher than the national baseline for conditions or behaviors which one wants to decrease or it is lower than the national baseline for conditions or behaviors one wants to increase.) In this case, the average annual percent change from the national baseline to the national year 2000 target can be multiplied by the number of years from the Washington baseline to the year 2000. The resulting percentage can be applied to the Washington baseline to establish the minimum change from baseline to target (outcome). The maximum change from baseline to target is the national standard. The panel recommends that these points be viewed as endpoints of a range of possible targets from which one point is selected. (See Attachment 2, Example 1 and Attachment 5, Calculation of Outcome Standard)*

   *For example, for the national health objective "reduce suicides in youth aged 15-19," the Healthy People 2000 target is 8.2 per 100,000, from a baseline of 10.3 per 100,000 in 1987. This is a decrease of 20.4% [(10.3-8.2)/10.3]. The decrease occurs over 13 years, yielding an average annual decrease of 1.6% (20.4%/13 years). The Washington baseline rate for 1989 through 1991 is 14.0 per 100,000. With nine years remaining to 2000, a comparable decrease for the state is 14.4% (1.6% per year for nine years), yielding possible target of 12.0 per 100,000 [12.0 is 14.4% lower than the baseline rate, i.e., 14.0 - (14.0*14.4%)]. This is the minimum change from baseline to target. The maximum change from baseline to target would be the U.S. target of 8.2 per 100,000. Thus, the endpoints for the range of possible targets are 8.2 to 12.0. The target of 11.2 per 100,000 chosen by the Technical Advisory Committee for the March PHIP Progress Report lies within this range and is, therefore, suitable.*

2. *The Washington baseline is better than the national baseline, but worse than the national year 2000 target. In these cases, it is recommended that the same method be applied as in situation 1, except that the derived point will now define the potential maximum change from baseline to target and the national year 2000 target will define the minimum change from baseline to target. Because it may not be biologically plausible or realistic for reasons other than biological plausibility to achieve the derived endpoint, it is characterized as a potential maximum change. If an outcome standard better than the national year 2000 target is selected, those who have chosen the standard must be able to document that their choice is realistic. One method of documentation would be to show that the rate has been achieved in other states, countries or sub-groups of the U.S. population. (See Attachment 2, Example 2)*

   For example, for the national health objective "reduce deaths from work-related injuries," the Healthy People 2000 target is 4 per 100,000 from a baseline five-year average of 5.9 per 100,000 in 1983 through 1987. This is a decrease of 32.2% over 13 years, which becomes 2.5% per year. The Washington baseline in 1991 is 4.7 per 100,000. With nine years remaining to 2000, a comparable decrease for the state is 22.3%, yielding a possible target of 3.7 per 100,000. The minimum change from baseline to target would be the US target of 4.0 per 100,000. Thus the endpoints for the range of possible targets are 3.7 to 4.0.

3. *The Washington baseline is better than the national baseline and better than the national year 2000 target. The approach recommended in situation 2 can be followed, except that the minimum change is to maintain the baseline rate. Documentation that the selection of a target is biologically plausible and/or realistic is necessary. (See Attachment 2, Example 3)*

The methods described above result in specifying a faster rate of change than the nation as a whole when Washington is doing worse than the nation and a slower rate of change when Washington is doing better. The methods describe how to generate a range of possible target values. The selection of a single point within that range depends on scientific and policy considerations such as: 1) are the factors which are causing Washington's rate to be worse than the national rate amenable to change, 2) how much of a priority is there at the national, state, and/or local level for addressing this condition, 3) is the difference between Washington's baseline and the national baseline an artifact of differential reporting, etc?

One tool to help in the selection of a point within the range is to apply the method described below for when there are no national targets. Using this method, one would calculate where one would expect to be in 2000 if the rate of change continues as it has over the past 5 to 10 years. A second tool to help decide which point to choose relates changes in health status to changes in risk factors. If there is a clear relationship between a risk factor and health outcome and if one can predict how much change will occur in the risk factor by 2000, then, based on that change, the amount of disease related to the risk factor (attributable risk), and the lag time between changes in risk factors and outcomes, one can predict the expected change in health status.

When there are no national year 2000 goals, the committee recommends that the Washington year 2000 target be based on a projected rate of change from historical data. In most instances, this means fitting a regression line to the historical data and extrapolating to 2000. The point can then be adjusted (raised or lowered) depending on the preventability of the condition, the political will for improvement, etc.

When there are no baseline data, data cannot inform the process for selecting an outcome standard. In these instances, the panel recommends that data be collected so that a *baseline can be measured before targets are specified and before interventions are initiated. Generally, there is no harm in adopting a national target or a subjectively identified target even though the target may be unrealistic. However, beginning an intervention program before measuring a baseline, can make it impossible to assess whether the program is having the desired impact. Once a baseline is developed, one could use the methods described above to develop an outcome standard.*

Thresholds

The fourth definition of threshold in Webster's II New Riverside University Dictionary (New Riverside Publishing Co., 1988) is "the intensity below which a ... stimulus ... can produce no response." Using this definition, a response is produced when a threshold is exceeded. The panel defined threshold standard as data which produce a response. The panel stresses that the initial response to exceeding a threshold is not intervention, but rather a closer look at the situation to determine what may be occurring. Additionally, a threshold is a way of measuring if one is progressing toward a goal at reasonable rate.

Given this latter way of looking at threshold standards, the first step toward measuring whether a threshold has been exceeded is to determine a reasonable rate of progression toward a goal. The most straightforward approach toward measuring expected progression is linear interpolation between the baseline data point and the outcome standard. Figure 2 in Attachment 5 illustrates where one would be expected to be each year to reach by 2000 the goal of 2918.6 hospitalizations for hip fractures per 100,000 women age 85 years and older.

The panel distinguishes between two types of thresholds. Thresholds may be trend based or group based. A trend based threshold compares data for a given year(s) to an expected or desired value for that year. The populations from which the two data points are derived are essentially the same. Group based thresholds compare data for similar time periods from a sub-group, such as a racial/ethnic group or a county, to a larger group.

Trend based thresholds

To ascertain whether a threshold has been exceeded for a given year, the panel recommends testing whether an actual data point differs from an expected data point. The expected data point is the point for that year on the interpolated line described above. The same statistical test described in the section on baselines can be used to determine whether the two points differ. If they do differ, one would conclude that a threshold had been exceeded and the situation needed to be studied more intensively from both data and programmatic perspectives to ascertain why the threshold was exceeded.

Non-overlapping 95% CIs around the actual and expected point estimates approximate the statistical test described above. While this approach is acceptable under any conditions, the panel recommends this approach, using Poisson CIs, when rates are less than 10 per 100 or when the number of events is less than 100, since there is difficulty in using a normal approximation to the binomial distribution under these conditions.

Group based thresholds

In some instances, one will want to know whether the rates for sub-groups exceed a threshold. Most often the sub-groups will be specific racial or ethnic groups or groups in relatively small geo-political areas, such as counties. This is conceptually the same as asking whether the rate for a particular group is significantly different from the rate of the population as a whole.

The same types of statistical tests that are used with trend based thresholds can be applied. However, with group based thresholds, the variance in the smaller group will be large compared to that of the total population. Therefore, as a short cut, it is possible to calculate a 95% CI only for the sub-group point estimate. If the rate from the larger group does not fall between the upper and lower 95% CI limits for the smaller group and the direction of the non-overlap indicates that the situation is worse in the smaller group, a threshold has been exceeded.

The panel cautions that one must be circumspect when comparing a sub-group to a population as a whole. Statistical inference is based on an assumption of independence of events. When a smaller group which is part of a larger group is compared to the larger group as a whole, the assumption of independence is violated. The larger the sub-population, the more the rate for the sub-population influences the overall rate and the less the independence. Those with data expertise must use their judgement to determine whether comparing a sub-group to a larger group is tenable.

As a general rule, if the sub-group comprises more than 20% of the larger group, the panel recommends not comparing the sub-group to the larger group. In these instances, one could calculate the rate for the larger group after subtracting the sub-group from both the denominator and numerator. The two groups would then be independent of one another and, therefore, they could be compared. Alternatively, a similar group from a different population could be sought for comparison purposes. For instance, it may be more informative to compare rates in King County to rates from other counties with relatively large urban centers rather than comparing King County to the rest of Washington.

If the sub-group is between 10% and 19% of the larger group, the data analyst could analyze the data both with the sub-group being part of the larger group and after subtracting the sub-group from the larger group. If there are differences in statistical inference using the two methods, the sub-group is too large to be included in the total group.

Additional considerations

As was discussed in the section on baselines, when there are a relatively small number of events, statistical procedures may not have the power to detect a real difference in rates. Conversely, statistics can show a difference between rates by chance, i.e. when there is no real difference. This latter situation is particularly likely when numerous statistical tests are performed. Therefore, the results of statistical testing and/or setting CIs must be interpreted with caution.

In most instances, the panel recommends that similar methods be used to calculate the baseline and threshold. Thus, if a three-year rolling average was used to calculate the baseline, a three-year average needs to be used to calculate the threshold. For group based thresholds, however, a single year could be compared to a three year average, if the single year was the mid-year of the average. For example, rates from a county based on data from 1990 to 1992 could reasonably be compared to 1991 Washington data. Judgement must be used to assure that the comparisons are tenable.

The methods described for ascertaining whether a threshold has been exceeded can only be used when a baseline and outcome standard have been developed. Additionally, these methods do not apply to items defined as sentinel events. With sentinel indicators, one event may be cause to investigate further.

Because this methodology relies on an evaluation of the variance of a given data point, it does not lend itself to setting one threshold standard against which the state can compare itself over time or against which communities or sub-populations can be evaluated. The panel discussed several procedures which could be used to specify threshold values for defined populations. The panel will work on refining these procedures for future versions of the PHIP.

Confidence intervals

The panel recommends the use of 95% CIs to aid in the determination of when a single year of data can be used as a baseline and to ascertain whether a threshold has been exceeded. The following recommendations are included to aid with the development of 95% CIs in a standardized manner.

If there are 100 or fewer events (i.e., the numerator is <= 100) or if the rate is less than 10% (e.g., 10,000 per 100,000), it is recommended that 95% CIs be calculated based on the Poisson distribution. Since Poisson tables are copyrighted, the panel cannot distribute them. If they are not readily available, the panel recommends using the method of Ury and Wiggins For 30 or more events, the Ury-Wiggins formula equals the exact Poisson 95% CIs to three significant figures. For 6 to 29 events the accuracy is to two significant figures. Since the panel recommends not calculating rates when the number of events is less than 5, the use of Ury/Wiggins' formulas is satisfactory. The Ury/Wiggins formulas for 95% CIs are as follows:

n = the number of events

Lower limit: $n - (1.96*\sqrt{n}) + 1.0$	
Upper limit: $n + (1.96*\sqrt{n}) + 2.1$	when $0 < n <= 50$
Upper limit: $n + (1.96*\sqrt{n}) + 2.0$	when $n > 50$

*These formulas specify the upper and lower CIs for the number of events. The number of events must be converted to a rate by dividing by the appropriate denominator and multiplying by the appropriate standard, e.g. 100,000 for rates expressed per 100,000. For example, if there were 20 events in a population of 40,000 the rate is 50.0 per 100,000 [(20/40,000)*100,000]. The lower and upper limits for 20 events are 12.23 [20 - (1.96*$\sqrt{20}$) + 1] and 30.87 [20 + (1.96*$\sqrt{20}$) + 2.1]. Conversion of these numbers to rates gives 30.6 [(12.23/40,000)*100,000] and 77.2 [(30.87/40,000)*100,000] as the lower and upper 95% CI limits around the rate of 50.0 per 100,000. Attachment 2 has been included to facilitate the process of finding the upper and lower number of events.*

If the number of events (numerator) is greater than 100 and if the rate is greater than 10%, it is acceptable for 95% CIs to be calculated based on the normal approximation to the binomial distribution. The standard formula for 95% CIs based on the normal approximation to the binomial distribution is

$$p \pm 1.96*\sqrt{p(1-p)/n} \quad ,$$

where p = the proportion of the population or sub-group with the characteristic (i.e. the rate expressed as the number of events divided by the number of individuals in the relevant population) and n = the number of individuals in the population or sub-group (i.e. the denominator used in calculating the rate).

There may be times when the use of a single method of developing CIs is expeditious (e.g., one wants to build a formula for CIs into a spreadsheet). In these instances, the panel recommends using the Ury/Wiggins formula, since for large numbers, the Poisson distribution is similar to the binomial distribution. It is also possible to use a computer program to calculate exact CIs based on the binomial distribution, rather than using Poisson or normal approximations to the binomial distribution.

Guidelines for the presentation of data

To interpret Washington data, it is helpful to place that data into a larger perspective. Therefore, if possible, data tables (or text, if there are no tables) should include a baseline and target (outcome) for Washington and comparable baseline and target data for the United States. To avoid confusion, data tables need to clearly differentiate Washington and national year 2000 targets.

The panel recommends that the following information be presented in the data tables or in the text if there are no data tables. If the information does not fit conveniently into the table itself, it can be included as a footnote to the table. Generally, the comments in the tables or footnotes can be short, since more detailed information is presented as appendices to these guidelines.

1. *If the baseline is not a single year or average, the method used to calculate the baseline should be noted. Generally, this means that it should be noted when the regression method has been used.*

2. *The table (or text) needs to note which years were used to generate the baseline and the year(s) to which the baseline data apply. When a single year or average is used, these two figures are the same.*

3. *Data sources and precise definitions for the number of events (i.e., the numerator if a rate is presented) need to be specified.*

4. *Data sources and precise definitions of the population to which the data refer (i.e., the denominator if a rate is presented) need to be specified.*

5. *To avoid ambiguity, the panel recommends that for children, age be presented in months, rather than years, e.g. does immunization status for those age two and under refer to those up to 24 months or those up to 36 months?*

Attachment 4 contains a sample table. Attachment 5 contains an example for the calculation of baselines, outcomes and thresholds for hospitalization for hip fracture.

Attachment 1
Summary of the PHIP data guidelines

This is a summary version of the data guidelines prepared by the PHIP Data Advisory Panel. These are recommendations, not hard and fast rules. For more detail, see the full report of the panel.

Rates and numbers

For most quantitative health status and risk indicators, express the baseline, outcome standard, and threshold standard as rates (i.e. number of events per population).

Do not use rates when:

- *There are fewer than five events, or;*
- *Small numbers of events serve as sentinel (red flag) indicators of public health concern (for example cases of polio).*

Use population denominators provided by the Washington State Center for Health Statistics.

If there is a comparable age-adjusted national indicator, age-adjust Washington data in the same manner as the national indicator. If adjustment is appropriate, adjust to the 1940 U.S. standard population, except for cancer data. Adjust cancer incidence to the 1970 U.S. standard population. Adjust cancer mortality to both the 1940 and 1970 standards.

In addition to rates (either crude or adjusted), also present the number of events.

Baseline calculation

Use the single most recent year of data for the baseline statistic when:

- *There are sufficient years of historic data (generally five or more) to permit the detection of a trend, and;*
- *The historic data indicate an underlying trend without substantial annual fluctuation. (The full report recommends both quantitative and qualitative ways of determining whether there is substantial annual fluctuation.)*

If there is substantial annual fluctuation using single years, the first recommended alternate method is to construct three-year moving averages, or three-year simple averages when there are at least 15 years of historic data. If this method reduces the fluctuation between adjacent data points sufficiently, use the most recent three-year average as the baseline.

If substantial fluctuations between adjacent data points persist after application of the averaging method, the second recommended method is fitting a regression line to historic data and using the end point of that line as the baseline statistic.

Outcome standards

Outcome standards are generally targets for the year 2000.

When national baseline data and year 2000 targets are available, there are three possibilities:

1. *Washington's baseline is worse than the national baseline. Determine the average annual percent change from the national baseline to the national year 2000 target, multiply it by the number of years from the Washington baseline to the year 2000, and apply this percentage to the Washington baseline. This becomes the minimum change from baseline to target. Set the state target somewhere between the minimum and a maximum of the national target. (See Attachment 3, Example 1 and Attachment 5, Calculation of Outcome Standard)*

2. *Washington's baseline is better than the national baseline, but worse than the national year 2000 target. Use the same method as above, but the derived point becomes the maximum change and the national target is the minimum change. (See Attachment 3, Example 2)*

3. *Washington's baseline is better than the both the national baseline and the national year 2000 target. Use the same method as above, except that the minimum change is to maintain the baseline rate. (See Attachment 3, Example 3)*

If the selected target is more ambitious than the national year 2000 target, be prepared to document that the choice is realistic from a scientific, social, or political point of view.

When there are no national year 2000 goals, base the Washington year 2000 target on a projected rate of change from historical data. Fit a regression line to historic data, extrapolate to 2000, and adjust the point up or down based on such considerations as preventability of the condition and the political will for improvement. When there are no historic data from which to develop a baseline, data cannot inform the process for selecting outcome standards.

Do not establish separate outcome standards for different racial/ethnic groups unless a condition is biologically linked to a specific group.

Threshold standards

There are two types of thresholds: trend-based and group-based.

Trend-based thresholds (same population, different years).

> *This type of threshold is a measure of progress toward a target, over time, for a given population. It compares data for a given year to an expected or desired value for that year. The most straightforward way to establish the expected or desired value is linear interpolation between the baseline data point and the outcome standard. This gives a series of expected data points for each intervening year. To determine if a threshold has been exceeded for a given year, test whether the actual data point differs from the expected data point.*

Group-based thresholds (same year, different populations)

> *This type of threshold measures how one group is doing compared to one or more other groups during the same time period. Usually it measures whether a rate for a sub-group is significantly different from the rate for the population as a whole. The sub-groups are often specific racial or ethnic groups. They may also be the entire populations of relatively small geo-political areas such as counties.*
>
> *In general, sub-groups comprising more than 20% of a larger group should not be compared to the larger group. Sub-groups comprising between 10% and 19% of the larger group may be analyzed both with the sub-group as part of the larger group and after subtracting the sub-group from the larger group. If there are differences in statistical inference using the two methods, the sub-group is too large to be included in the total group.*

The full report has more detail on how to establish thresholds and determine whether they have been exceeded. There is also a detailed discussion of how to calculate and use confidence intervals.

Datapresentation

Data tables (or text if there is no table) should include the following:

- *Baseline and target (outcome standard) for Washington and comparable baseline and target for the United States.*
- *Year or years used to produce baseline data. If the regression method was used, indicate what years of historic data were used and for which year the baseline applies.*
- *Data sources and precise definitions for the number of events (numerator) and the population (denominator).*
- *Notes regarding such factors as whether rates are for incidence or prevalence and whether death rates are age-adjusted.*

Present age for very young children in months to avoid confusion about whether a term such as "age two and under" means those up to 24 months or those up to 36 months.

See Attachment 4 for sample table.

Attachment 2
Poisson 95% confidence intervals

n	LL	UL	n	LL	UL	n	LL	UL
1	0.04	5.06	34	23.57	47.53	67	51.96	85.04
2	0.23	6.87	35	24.40	48.70	68	52.84	86.16
3	0.61	8.49	36	25.24	49.86	69	53.72	87.28
4	1.08	10.02	37	26.08	51.02	70	54.60	88.40
5	1.62	11.48	38	26.92	52.18	71	55.48	89.52
6	2.20	12.90	39	27.76	53.34	72	56.37	90.63
7	2.81	14.29	40	28.60	54.50	73	57.25	91.75
8	3.46	15.64	41	29.45	55.65	74	58.14	92.86
9	4.12	16.98	42	30.30	56.80	75	59.03	93.97
10	4.80	18.30	43	31.15	57.95	76	59.91	95.09
11	5.50	19.60	44	32.00	59.10	77	60.80	96.20
12	6.21	20.89	45	32.85	60.25	78	61.69	97.31
13	6.93	22.17	46	33.71	61.39	79	62.58	98.42
14	7.67	23.43	47	34.56	62.54	80	63.47	99.53
15	8.41	24.69	48	35.42	63.68	81	64.36	100.64
16	9.16	25.94	49	36.28	64.82	82	65.25	101.75
17	9.92	27.18	50	37.14	65.96	83	66.14	102.86
18	10.68	28.42	51	38.00	67.00	84	67.04	103.96
19	11.46	29.64	52	38.87	68.13	85	67.93	105.07
20	12.23	30.87	53	39.73	69.27	86	68.82	106.18
21	13.02	32.08	54	40.60	70.40	87	69.72	107.28
22	13.81	33.29	55	41.46	71.54	88	70.61	108.39
23	14.60	34.50	56	42.33	72.67	89	71.51	109.49
24	15.40	35.70	57	43.20	73.80	90	72.41	110.59
25	16.20	36.90	58	44.07	74.93	91	73.30	111.70
26	17.01	38.09	59	44.94	76.06	92	74.20	112.80
27	17.82	39.28	60	45.82	77.18	93	75.10	113.90
28	18.63	40.47	61	46.69	78.31	94	76.00	115.00
29	19.45	41.65	62	47.57	79.43	95	76.90	116.10
30	20.26	42.84	63	48.44	80.56	96	77.80	117.20
31	21.09	44.01	64	49.32	81.68	97	78.70	118.30
32	21.91	45.19	65	50.20	82.80	98	79.60	119.40
33	22.74	46.36	66	51.08	83.92	99	80.50	120.50
						100	81.40	121.60

To calculate a rate: use the number of events (n) as the numerator, the population of the area as the denominator. Rates are usually expressed in units of "per 100,000," so multiply the result by 100,000 to obtain the rate.

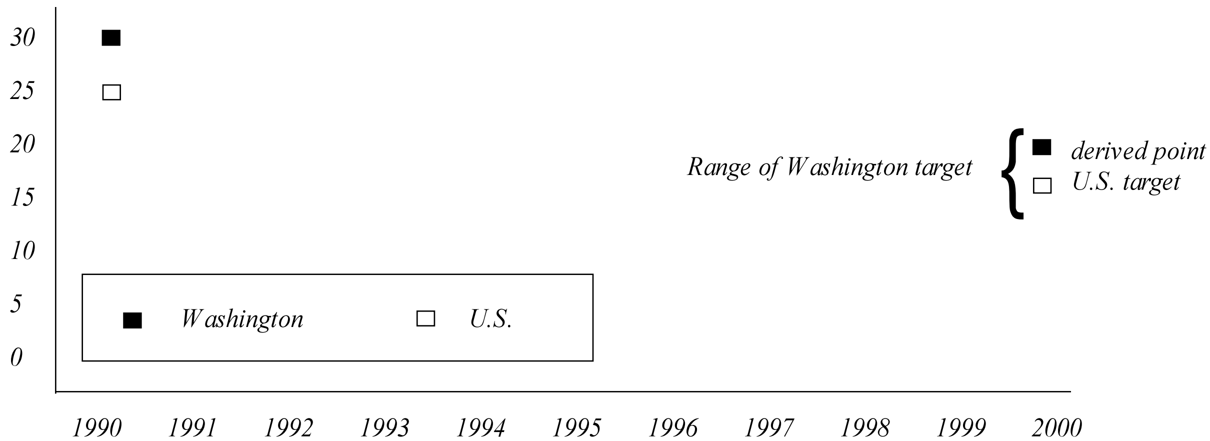
To calculate 95% Confidence Interval: repeat the rate calculation using the lower limit (LL) and the upper limit (UL).

C.I. Method: Ury HK, Wiggins AD. Am J Epidemiol 1985; 122(1):197-8

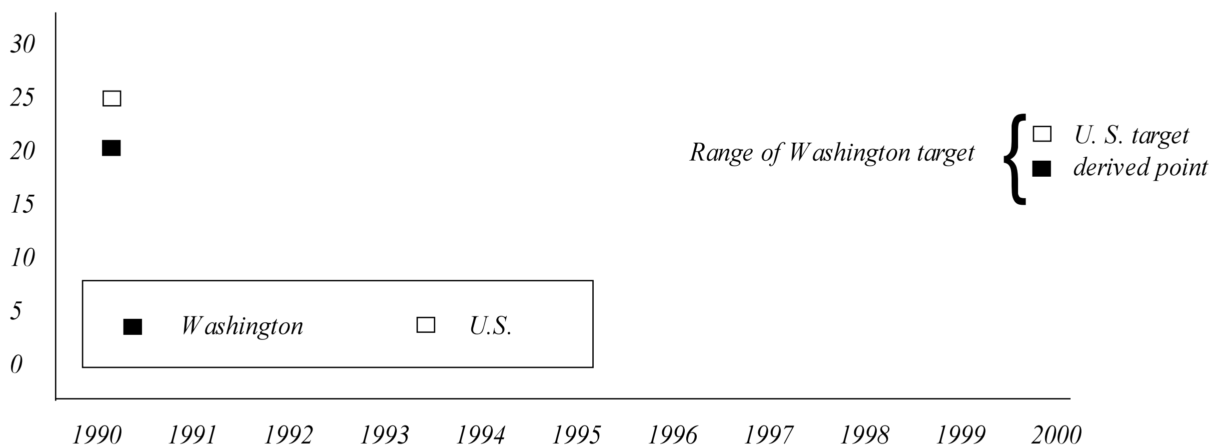
Attachment 3

Ranges for outcome standards

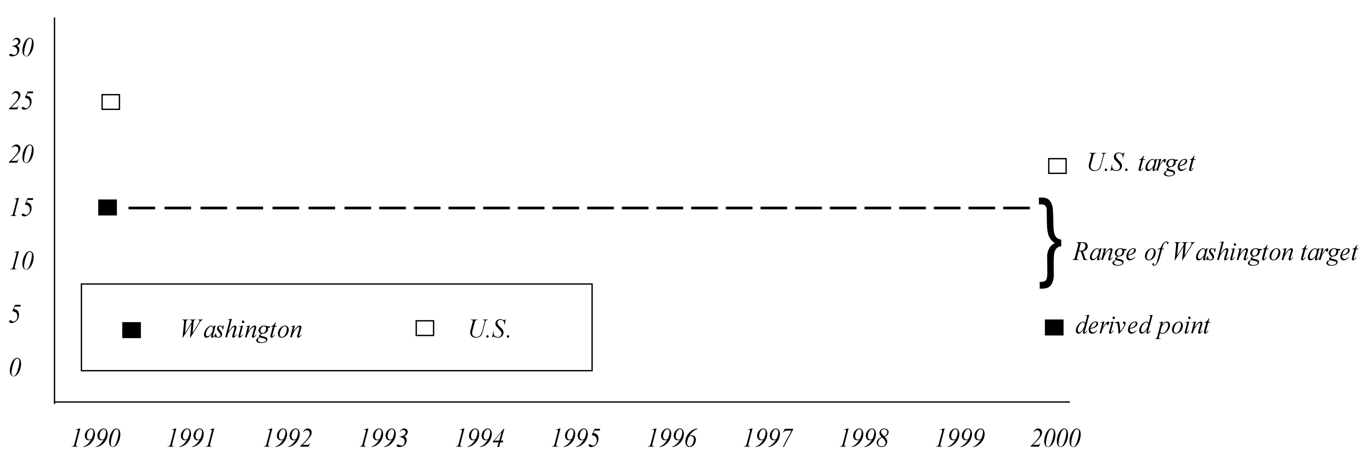
Example 1



Example 2



Example 3



Attachment 4
Sample table

Standards for Falls and Fall-Related Injuries

	Washington State				United States		
	Baseline			Year 2000	Baseline		Year 2000
	Year(s)	Count	Rate	Target Rate	Year(s)	Rate	Target Rate
Fall Deaths							
Total Population	1992	302	3.3	3.0	1987	2.7	2.3
Age 65-84	1990-92	296	18.7	16.5	1987	18.0	14.4
Age 85+	1990-92	304	174.6	153.6	1987	131.2	105.0
Hospitalization for Hip Fracture							
Age 65+	1990-92	12,297	700.6	630.6	1988	714.0	607.0
Women Age 85+*	1990-92	3812	3074.7	2224.9	1988	2721.0	2177.0

Data Source(s):

Deaths - Vital Statistics

Hospitalizations - CHARS

Population statistics - Dept. of Health, Center for Health Statistics, 7/5/94

Case Definition(s):

Falls and fall-related injuries include all ddeaths coded to E880-E888.

Hip fracture includes all resident hospitalizations with a principal diagnosis of N820.

Additional Notes:

Rates are per 100,000 resident population.

Death rates for the total population are age-adjusted.

* U.S. baseline and target are for white women 85 and older.

Attachment 5
Annotated example using hospitalization for hip fracture

The Violence and Injury sub-section of the Key Priority Public Health Problems Section of the Public Health Improvement Plan includes standards for falls and fall-related injuries. One of the measures in this sub-section is hospitalizations for hip fracture. Data for hospitalizations for hip fracture are available through the Comprehensive Hospital Abstract Reporting System (CHARS). The Washington State Injury Prevention Program (WSIPP) uses an enhanced CHARS data set which begins with 1989 data. At the time data were requested for the PHIP, WSIPP had CHARS data for 1989 through 1992.

Since Healthy People 2000 (U.S. Public Health Service, DHHS Publication No. (PHS) 91-50212, 1991) includes standards for hospitalization for hip fracture, data for this indicator in the PHIP are developed so that they are comparable to the data in Healthy People 2000. Specifically,

1. *a case is defined as a Washington resident with a hospital discharge International Classification of Diseases, 9th Revision, Clinical Modification (ICD9-CM) code of 820 as the primary diagnosis;*
2. *hospital discharges, not people, are counted (i.e., a person hospitalized twice for the same event is counted twice); and*
3. *the same age and sex groupings are used as in Healthy People 2000.*

For women aged 85 years and older, Healthy People 2000 specifies white women only. Because CHARS does not include race, the Washington data for this age group are not strictly comparable to the data in Healthy People 2000. In Washington, less than 3.0% of women ages 85 years and older are non-white. While this is a relatively small percentage, this difference in data development must be borne in mind when comparing Washington data for this indicator to the national data.

Calculation of baseline

*Tables 1 provides the number of events, the population, rates and 95% Poisson confidence intervals (CIs) on which the analysis is based. For men and women ages 65 years and older, the 1989 rate of hospital discharge is 680 per 100,000 (3783/556077*100000). Because the rates in this example are less than 10%, 95% Poisson CIs are calculated. The upper and lower 95% Poisson CI are calculated from Ury and Wiggins (AJE 122:197-198,1995) formulas for the lower and upper limits, n-(1.96*šn)+1 and n+(1.96*šn)+2, where n is the number of events. Thus, for the lower limit, n-(1.96*šn)+1 = 3783-(1.96*61.5)+1 = 3663 events. The number of events converts to a rate of 659 per 100,000 (3663/556077*100000).*

Table 1: Hospital discharges for hip fracture (ICD9-CM 820)

	Men and Women Ages 65+				Women age 85+			
Year	Number	Population	Rate	95% CIs	Number	Population	Rate	95% CIs
1989	3783	556077	680	659-702	1153	38070	3029	2856-3209
1990	3855	571404	675	654-696	1190	39560	3008	2840-3184
1991	4125	585717	704	683-726	1297	41316	3139	2971-3315
1992	4317	598102	722	700-744	1325	43102	3074	2911-3244
1990-92	12297	1755223	701		3812	123978	3075	

Figures 1 and 2 provide the rates and 95% CIs graphically. Both figures show wide overlapping CIs for adjacent years, indicating that rates may be stable enough to use the most recent year as the baseline. However, with only four years of data, it is difficult to determine whether the trend of increasing hospitalizations shown in Figure 1 is real and the 1992 rate is a good representation of where we are at baseline. Given this difficulty, the staff of WSIPP decided to use a three year average for the baseline rate for hip fracture hospitalization for people ages 65 years and older.

Figure 1
Hospitalizations for Hip Fractures
Men and Women, Ages 65 Years and Older

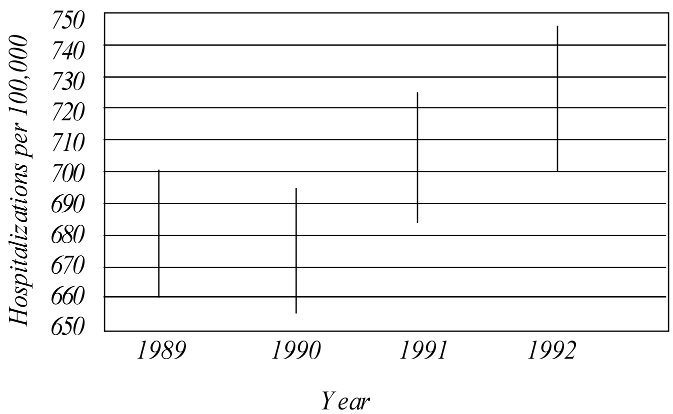
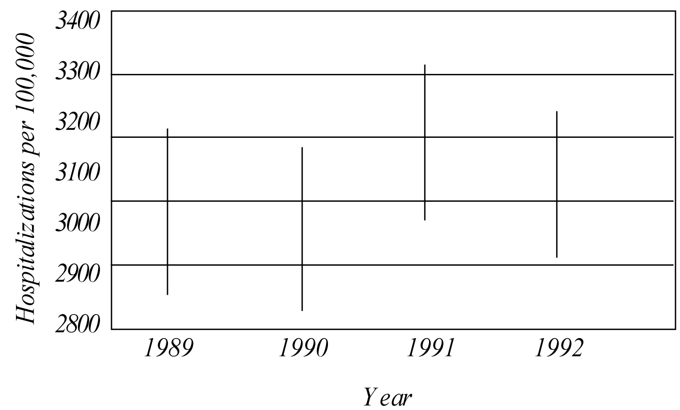


Figure 2
Hospitalizations for Hip Fractures
Women, Ages 65 Years and Older



For hip fracture hospitalization in women ages 85 years and older, the 1992 rate is very similar to the three-year average rate from 1990 to 1992, indicating that 1992 may be a reasonably accurate representation of where we are at baseline. However, for consistency with the baseline for ages 65 years and older, WSIPP staff decided to use a three-year average for this baseline also.

The three-year average is derived by adding the number of hospital discharges for 1990, 1991 and 1992 and dividing by the number of person-years for the same period. The number of person-years is calculated by adding the population for 1990, 1991 and 1992.

Calculation of outcome standard

Since there are national year 2000 targets for these indicators, WSIPP assessed the expected change in the indicators from baseline to target at the national level. Table 2 provides the data necessary for these analyses.

Table 2: Data for calculation of Washington targets

	Healthy People 2000				Washington	
Group	Baseline		Target		Baseline	
	Year	Rate	Year	Rate	Year	Rate
Ages 65+	1988	714	2000	607	1990-1992	701
White Women, Ages 85+	1988	2721	2000	2177	1990-1992	3075

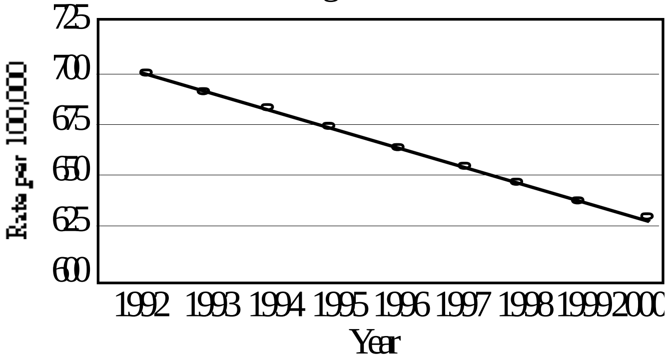
For ages 65 years and older, the percent change from the national baseline to target is approximately 15% [(714-607)/607*100]. Since the baseline was calculated in 1988 and the goal is for 2000, there are 12 years in which to achieve the 15% decrease, becoming an average annual decrease of 1.25% (15%/12). Washington's baseline covers a three year period from 1990 to 1992, leaving eight years until 2000. A decrease of 1.25% per year for eight years is a total decrease of approximately 10% (1.25%*8). This decrease is applied to the baseline (700.6 per 100,000) to arrive at a potential target of 631 per 100,000 [701-(10%*701)]. In this example, the calculated target represents the minimum change from baseline to target. The maximum change is to use the national target. Given that historical data indicates that rates may be increasing and the expectation that the level of funding for health promotion programs for the elderly will remain stable at best, WSIPP staff selected the calculated value (i.e., the minimum change) for the year 2000 target.*

*For women ages 85 years and older, the percent change from baseline to target is approximately 20% [(2721-2177)/2721*100] over a 12 year period. This is an average annual decrease of 1.67% (20%/12). A decrease of 1.67% per year for eight years is a total decrease of 13.3% (1.67%*8) which becomes a potential target of 2665 per 100,000 [3075-(13.3%*3075)]. Again, this figure represents the minimum change from baseline to target, with the national baseline representing the maximum change. In deciding which point within this range to use as the year 2000 target, WSIPP staff considered that 1) at current levels of funding and intervention, the rate of hospital discharge in this group seems to be stable; 2) the levels of funding and intervention are likely to remain stable, at best; and 3) since the national data refer to white women and the Washington data include all women, the national target may be artifactually low for Washington. The WSIPP staff selected the minimum change as the target value.*

Threshold calculation

Figure 3 illustrates the anticipated progress toward the year 2000 target for hospital discharges for hip fracture in men and women ages 65 years and older. If we are progressing toward our goal as expected, by 1993, we expected the rate of hospital discharges for hip fracture to be 692 per 100,000. Since the data for 1993 are not available, let us assume that in 1993 there are 4500 hospital discharges for hip fracture among 600,000 people ages 65 years and older, yielding a rate of 750 per 100,000. Have we exceeded a threshold?

Figure 3
Targets for Hospitalization for Hip Fracture, Ages 65 Years and Older



*Since the baseline was based on only four years of data, the three-year average can be perceived as a simple or moving average. However, since the data guidelines specify creating simple averages only when there are approximately 15 years of historical, it may be best to conceive of this as a moving average. If this is the case, the rate for 1993 should be created as the next point in that moving average. Thus, the number of hospital discharges for 1991 to 1993 are divided by the number of person-years for the same time period and multiplied by 100,000, yielding a rate of 726 per 100,000 (12942/ 1783819*100000). The 95% lower Poisson CI is 713 per 100,000 [((12942-(1.96** $\sqrt{12942}$*)+1)/1783819*100000], which is higher than the interim target of 692 per 100,000. This indicates that we may not be progressing toward the target as expected. The next step is to determine why we do not seem to be progressing as planned and based on that determination decide whether increased efforts are necessary and possible.*